BioMistral: A Collection of Open-Source Pretrained Large Language Models for Medical Domains

Abstract

Large Language Models (LLMs) have demonstrated remarkable versatility in recent years, offering potential applications across specialized domains such as healthcare and medicine. Despite the availability of various open-source LLMs tailored for health contexts, adapting general-purpose LLMs to the medical domain presents significant challenges. In this paper, we introduce BioMistral, an open-source LLM tailored for the biomedical domain, utilizing Mistral as its foundation model and further pre-trained on PubMed Central. We conduct a comprehensive evaluation of BioMistral on a benchmark comprising 10 established medical question-answering (QA) tasks in English. We also explore lightweight models obtained through quantization and model merging approaches. Our results demonstrate BioMistral's superior performance compared to existing open-source medical models and its competitive edge against proprietary counterparts. Finally, to address the limited availability of data beyond English and to assess the multilingual generalization of medical LLMs, we automatically translated and evaluated this benchmark into 7 other languages. This marks the first large-scale multilingual evaluation of LLMs in the medical domain. Datasets, multilingual evaluation benchmarks, scripts, and all the models obtained during our experiments are freely released.